

# How prior knowledge affects selective attention during category learning: An eyetracking study

ShinWoo Kim · Bob Rehder

Published online: 2 December 2010  
© Psychonomic Society, Inc. 2010

**Abstract** Research has shown that category learning is affected by (a) attention, which selects which aspects of stimuli are available for further processing, and (b) the existing semantic knowledge that learners bring to the task. However, little is known about how knowledge affects what is attended. Using eyetracking, we found that (a) knowledge indeed changes what features are attended, with knowledge-relevant features being fixated more often than irrelevant ones, (b) this effect was not due to an initial attentional bias toward relevant dimensions but rather emerged gradually as a result of observing category members, and (c) this effect grew even after a learning criterion was reached, that is, despite the absence of negative feedback. We argue that models of knowledge-based learning will remain incomplete until they specify mechanisms that dynamically select prior knowledge in response to observed category members and which then directs attention to knowledge-relevant dimensions and away from irrelevant ones.

**Keywords** Category learning · Selective attention · Eyetracking · Categorization · Thematic knowledge

## Introduction

Selective attention has played a prominent role in theories of category learning ever since the finding that learning difficulty correlates with the number of stimulus dimensions needed for classification (Shepard, Hovland, & Jenkins, 1961). In both exemplar (Medin & Schaffer, 1978; Nosofsky, 1986) and prototype (Hampton, 1995; Nosofsky, 1992; Smith & Minda, 1998) models, selective attention is formalized in terms of weights that different stimulus dimensions have on classification. Rule-based models also assume that learners selectively attend to the dimensions referred to by the current hypothesis being tested (Smith, Patalano, & Jonides, 1998; also see Maddox, 2002; Maddox & Dodd, 2003). Furthermore, some models include mechanisms specifying how selective attention changes with learning (Kruschke, 1992; Kruschke & Johansen, 1999; Nosofsky, Palmeri, & McKinley, 1994).

Another important finding is that the learning of categories is affected by whether the new category can be related to existing semantic knowledge (see Murphy, 2002, for a review). For example, in Murphy and Allopenna's seminal study (1994), subjects learned two categories in which features of each category could be related to a particular theme. Features of one category included "drives on glaciers," "made in Norway," and

---

S. Kim (✉)  
Department of Industrial Psychology,  
Kwangwoon University,  
Wolgae-Dong, 447-1,  
Nowon-Gu, Seoul, South Korea 139-701  
e-mail: shinwoo.kim@kw.ac.kr

B. Rehder  
Department of Psychology,  
New York University,  
New York, NY, USA

“heavily insulated” and those for another included “drives in jungles,” “made in Africa,” and “lightly insulated.” Subjects learned to distinguish these categories far faster—presumably because the features could be subsumed under the themes—as compared to categories whose features shared no theme (see also Rehder & Ross, 2001). Subjects also show better learning of theme-related features versus those that are unrelated (Heit & Bott, 2000; Kaplan & Murphy, 2000).

But although selective attention and prior knowledge both affect the learning of categories, little is understood about how knowledge affects attention. This question is important because any theory of how knowledge influences learning is incomplete without an account of how it alters what category information is attended and thus processed. For example, although category learning models such as Baywatch (Heit & Bott, 2000) and KRES (Rehder & Murphy, 2003; Harris & Rehder, 2006) account for effects of knowledge by assuming that it moderates basic associative learning mechanisms, neither model postulates changes to selective attention per se, an important omission given the importance of attention cited above. However, this omission is understandable given that virtually *nothing* is known about how knowledge alters what is attended: modeling is impossible if there is no data to model. Thus, to further our understanding of knowledge effects in particular, and category learning in general, a first step is to establish some basic empirical facts regarding how knowledge modulates attention. To this end, we used eyetracking as a relatively direct measure of selective attention during knowledge-based category learning.

#### How knowledge might affect attention

Our study is guided by three open questions regarding how prior knowledge might affect attention during learning<sup>1</sup>. The first question concerns whether knowledge induces *any* change to what is attended. On one hand, many theorists have suggested that knowledge exerts its effect by directing attention to some sources of information at the expense of others (Murphy & Medin, 1985; Pazzani, 1991; Wisniewski, 1995). For example, if a subset of category features can be

<sup>1</sup> We use the term “attention” to refer to the overt perceptual attention required to obtain information about features of a stimulus. This notion of attention should be distinguished from an alternative interpretation common in the category learning literature in which attention refers to the weight that a stimulus dimension has on the ultimate categorization decision (also referred to as “decision weights”). This distinction is important because, for example, although a categorizer may obtain feature information from many stimulus dimensions via overt (perceptual) attention, those dimensions may have unequal influence on how an object is classified.

related to a common theme, it is natural to suspect that thematic knowledge might direct attention to those features and away from others (see Kaplan & Murphy, 2000, for discussion). However, there is also ample reason to believe that knowledge affects learning not via attention but rather by changing how features are encoded. It is well known that memory depends on the “depth” or meaningfulness of the encoding process (e.g., Bower, Clark, Lesgold, & Winzens, 1969; Craik & Lockhart, 1972; Craik & Tulving, 1975), suggesting that knowledge might speed category learning simply because of the better feature memory it supports. In addition, knowledge might allow classification to become an act of inference based on semantic understanding of the category. Consistent with these possibilities, research has shown that models like Baywatch and KRES that incorporate encoding and inferential processes but not selective attention are nonetheless sufficient to account for numerous effects of knowledge on category learning. Thus, our first goal is to determine whether knowledge indeed has any effect on selective attention.

Assuming that knowledge affects attention, a second question concerns the time course of that effect. Some theorists suggest that the role of knowledge is to *preselect* dimensions (or hypotheses) for further testing (Keil, 1981). For example, Pazzani’s (1991) rule-based PostHoc model initially tests hypotheses involving features that are relevant to goals (knowledge) associated with the category. Kruschke (1993) suggested that his ALCOVE model can account for prior knowledge by setting *initial* attention weights on the related dimensions higher than on the unrelated ones. Consistent with this proposal, some studies have reported an effect of knowledge that emerged very early in category learning (e.g., Heit, 1995; Kaplan & Murphy, 2000). Alternatively (or in addition), the effect of knowledge might increase during the course of learning as a result of observing category members. Because prior knowledge consists of representations in semantic memory, people may need to observe multiple category exemplars in order for a common theme to become sufficiently active in memory and thus noticed (see Heit & Bott, 2000, for discussion). In addition, a learner may only begin to make use of knowledge when a simpler strategy (e.g., one-dimensional rule-testing) fails to yield an acceptable solution. Thus, our second question is whether any effect of knowledge on attention is limited to which sources of information learners initially consider or whether it can emerge as a result of experience with category members.

Assuming that attention shifts as a result of observing category members, a third question concerns whether error feedback is required to mediate those shifts. Popular models that account for category learning in the absence of prior knowledge assume that learners respond to error feedback

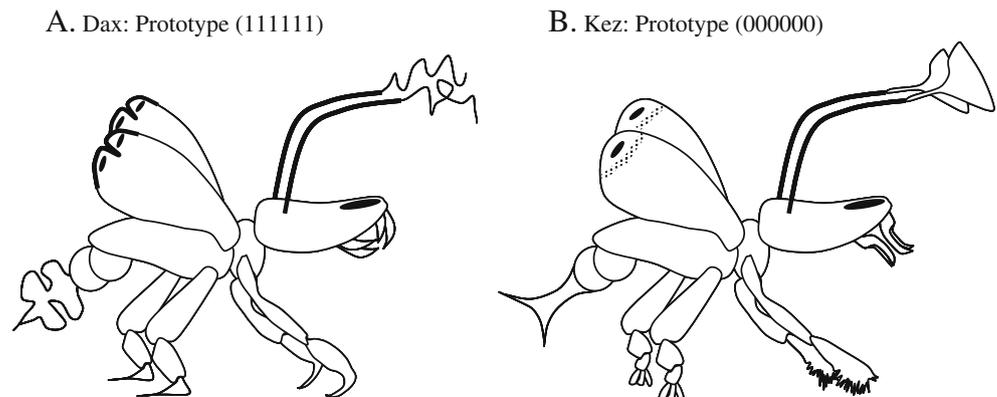
by shifting attention to stimulus dimensions that will reduce error in the future (e.g., Kruschke, 1992, 2001). This error-driven account might be extended to thematic category learning, because, for example, negative feedback might serve as a cue indicating to the learner to use prior knowledge (or to use *different* knowledge), which in turn might induce a shift in attention to features related to that knowledge. Alternatively, shifts in attention may also arise in the absence of error. That prior knowledge affects how subjects spontaneously sort exemplars into categories (and possibly which features are attended) in category construction tasks in which corrective feedback is absent (e.g., Kaplan & Murphy, 1999; Medin, Wattenmaker, & Hampson, 1987; Spalding & Murphy, 1996) suggests that attention may shift even on those trials during a supervised learning task in which error is absent (Blair, Watson, & Meier, 2009a).

To address these three questions, we conducted an eyetracking study of supervised category learning in which a subset of each category features could be related to a common theme. In cognitive research, eyetracking has been proven to be an effective tool to study on-line attention (e.g., Ferreira & Clifton, 1986; Haider & Frensch, 1999; Just & Carpenter, 1984; Lee & Anderson, 2001; Rayner, 1998). In recent years, it has been successfully applied to studying selective attention in category learning in the absence of knowledge (Blair et al., 2009a; Blair, Watson, Walshe, & Maj, 2009b; Rehder & Hoffman, 2005a, 2005b; Rehder, Colner, & Hoffman, 2009; Watson & Blair, 2008). We now use eyetracking to study how attention is affected by prior knowledge.

## Overview of experiments

We constructed two categories of ants labeled “Dax” and “Kez” from six binary dimensions. Figure 1 presents an example of the prototypes of the Dax and Kez categories. Unlike previous studies of thematic category learning using verbal feature descriptions, we used spatially separated pictorial features suitable for eyetracking.

**Fig. 1** Example prototypes of the Dax and Kez categories



Category exemplars were constructed using a one-away structure that included prototypes of each category (Table 1). In each category, four *related features* were associated with a theme by describing them as useful in either a cold or a hot climate. The other two *neutral features* were unrelated to these themes. Table 2 presents example feature descriptions for the prototypes in Fig. 1, where antenna, mouth, forearm, and foot are theme-related, and tail and wing are neutral. Participants learned these feature descriptions prior to training to determine how the themes of cold and hot climate would affect their subsequent learning. They were provided with no initial information regarding which feature went with which category.

Because our materials were novel and required participants to first associate each feature to its theme (also see Krascum & Andrews, 1998), in **Experiment 1** we first conducted a non-eyetracking study to establish whether this “prior knowledge” will induce standard knowledge effects. First, to confirm whether such knowledge allows faster learning, we compared learning performance in the *related condition* in which themes were present (as described above) to an *unrelated condition* in which themes were absent (i.e., all six dimensions were neutral to the themes). Second, in the related condition, we compared learning of the related dimensions to the neutral ones, to confirm that the former were learned better than the latter (as in, e.g., Heit & Bott, 2000). In **Experiment 2**, we conducted an eyetracking study to address the main questions surrounding attention and prior knowledge.

## Experiment 1

### Method

#### Materials

Dax and Kez categories were constructed from six binary dimensions: antenna, mouth, forearm, foot, tail, and wing.

**Table 1** Abstract structure for the Dax and Kez categories

Exemplars	Dimensions					
	Tail	Foot	Wing	Mouth	Forearm	Antenna
Dax						
D0	1	1	1	1	1	1
D1	1	1	1	1	1	0
D2	1	1	1	1	0	1
D3	1	1	1	0	1	1
D4	1	1	0	1	1	1
D5	1	0	1	1	1	1
D6	0	1	1	1	1	1
Kez						
K0	0	0	0	0	0	0
K1	0	0	0	0	0	1
K2	0	0	0	0	1	0
K3	0	0	0	1	0	0
K4	0	0	1	0	0	0
K5	0	1	0	0	0	0
K6	1	0	0	0	0	0

Table 1 presents a category structure in which the Dax and Kez prototypes were 111111 (Fig. 1a) and 000000 (Fig. 1b), respectively. Across subjects, there were four assignments of features to the Dax/Kez prototypes: 111111/000000, 101010/010101, 010101/101010, and 000000/111111. This balancing resulted in each feature being paired with each other and the two themes an equal number of times. In the related condition, Daxes were related to the cold, tundra-like theme and Kezes to the hot, desert-like theme. To relate categories to the themes, four of the six dimensions were accompanied with theme-related descriptions and the remaining two had neutral descriptions (see Table 2 for an example). To cancel out any effect of dimension's screen location (e.g., top vs. bottom) or type (e.g., head vs. tail), different physical dimensions instantiated the roles of the related and neutral dimensions. Across subjects, the neutral dimensions were either tail/foot, wing/mouth, or forearm/antenna, with the remainder being theme-related. In the unrelated condition, all dimensions were neutral. The Appendix presents the three types of descriptions (tundra, desert, or neutral) for each of the 12 features in Fig. 1. The two experimental conditions (related vs. unrelated), four

**Table 2** Example feature descriptions for the Dax and Kez prototypes in Fig. 1. Four related dimensions were associated with either a cold or a hot climate. The other two neutral dimensions were unrelated to these themes

Dimension	Dax [tundra/cold theme]	Kez [desert/hot theme]
Related		
Antenna	Because the temperature is very low, parts of ants' eyes (e.g., cornea, iris, pupil) often freeze and the ants become blind. When that happens, this thread type of flexible antennae is used to detect close objects.	Because the air is hot and dry, the ants are vulnerable to dehydration. To maintain hydration, the ants use this fan type of antennae to absorb water vapor from the air.
Mouth	Because the ground is frozen, the ants need to cut and break tough soil in search of their food. This type of mouth with sharp incisors serves this function.	Because sources of food are covered with sand, they need to be cleared before swallowing. The inner surface of the ants' mouth has short but stiff hairs that filter out these impurities.
Forearm	Because of frequent blizzards, the ants need to anchor themselves during high winds. This type of forearm allows the ant to hold its position.	Because the ants' preys (e.g., fleas) hide in sand, the ants use this type of forearm to sweep the sand and detect the prey.
Foot	Because the ground surface is slippery, the ants need to have wide feet to maintain their footing.	Because the ground surface is extremely hot, the ants switch the toe that comes into contact with the ground in each step to avoid burning.
Neutral		
Tail	The ants feed proteins stored in the humps to their larvae using the sharp nozzle in the end of tail.	The ants lay a large number of eggs at a time. This trumpet-shaped tail allows the ants to deliver a large number of eggs.
Wings	While flying, the ants control their rapid changes in direction by adjusting the fore- and rear-flaps in each wing.	The ants have red spots in the wing ends. The color becomes brighter in the mating season by the hormones produced in the gray area.

assignments of features to categories and three assignments of related/neutral dimensions resulted in 24 experimental cells.

### Participants

Participants were 30 New York University undergraduates who volunteered for course credit. They were randomly assigned to the 24 cells with the constraint of at least one person in one cell. This resulted in 14 and 16 participants in the related and unrelated conditions, respectively.

### Procedure

The experiment consisted of three phases: knowledge acquisition, category learning, and a single-feature test. During knowledge acquisition, participants studied 12 features. Each screen displayed an ant with one visible feature and the other five features were hidden behind gray rectangles (see Fig. 2, for an example). Below the ant were descriptions about the visible feature. Participants studied the 12 features at their own pace by navigating 12 screens with left/right arrow keys. The bottom of each screen displayed its number (1–12), and the presentation order of the features was randomized for each participant. At this point, no information was provided regarding which category a feature would be associated with during training.

To ensure learning, participants were required to take a multiple-choice test followed by a recall test. Both tests consisted of 12 questions, one for each feature. In the multiple-choice test, a question presented an ant with one visible feature, and participants chose one of the four alternatives (Fig. 2). The order of the questions was randomized for each subject. Immediate feedback was provided for each question, and after the test, the total

number of errors. When any error occurred during the test, participants were returned to the initial screens for additional study and then retook the test that presented only the questions they missed. This process repeated until all questions were answered correctly.

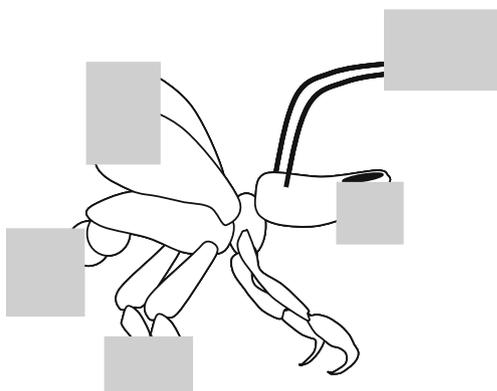
The recall test ensured that participants could not only recognize but also recall the feature descriptions during category learning (otherwise, there would be no effects of knowledge). This test was the same as the multiple-choice test except that participants verbally described each feature instead of making a choice. The experimenter provided feedback for each question, and after the test, the total number of errors. Any error during the recall test obligated the participant to restart the knowledge acquisition phase including initial learning, multiple-choice, and recall tests. This process repeated until participants answered all recall test questions correctly. The knowledge acquisition took about 12 minutes.

The category learning phase began with two practice trials. Training blocks then presented the 14 training exemplars (Table 1) in random order. Each trial began with a cross fixation (+) appearing for 1.8 s followed by presentation of an exemplar. Participants classified the exemplar by pressing “z” for Dax or “?” for Kez. Feedback was provided in words below the exemplar (“Correct” or “Wrong”) and the exemplar remained visible for 3.8 s after the response. For the practice trials, features were replaced with geometric shapes, and one trial displayed positive feedback and the other trial displayed negative feedback. Training ended after two errorless blocks in a row or after the 15th block. Participants were informed of how close they were to this goal after each block.

In the single-feature test, participants classified 12 features (as they did during training) randomly presented in each trial (as in Fig. 2). No feedback was provided. After each choice, participants rated confidence in the decision by positioning a slider on a scale whose left and right ends were labeled “Very Uncertain” and “Very Certain.” The slider could be set to 21 distinct positions, and responses were scaled to a range from 0 to 100. The whole experiment took about 50 minutes.

### Results

There were no effects of the counterbalancing factors in any of the following analyses, and thus the results are collapsed over these factors. Participants were very accurate in the tests during knowledge acquisition. No participant made more than a total of seven errors; 22 participants committed no errors. Related (0.97) and unrelated (0.98) participants



**Fig. 2** An example of a multiple-choice question

were both equally accurate ( $t < 1$ ), suggesting that the materials were easy to learn.

In category learning, 12 (of 14) related and 14 (of 16) unrelated participants reached the learning criterion of two consecutive errorless blocks. The related learners reached the criterion in fewer blocks (5.50) than the unrelated learners (8.57),  $t(24) = 2.85$ ,  $p < .01$ , while committing fewer total errors (8.67 vs. 19.07),  $t(24) = 3.29$ ,  $p < .01$ . These results replicated past findings (e.g., Murphy & Allopenna, 1994) and confirm effectiveness of our material in inducing the knowledge effect. Classification RTs decreased during training with no differences between the related versus unrelated conditions.

In the single-feature test, we asked whether related learners showed better learning of the related dimensions than the neutral ones (e.g., Heit & Bott, 2000). Table 3 indicates that the related learners were more accurate on the related dimensions (0.89) than the neutral ones (0.71),  $t(11) = 1.79$ ,  $p = .10$ . They also classified the related dimensions faster (2.9 s) than the neutral ones (4.1 s),  $t(11) = 1.48$ ,  $p = .17$ . Although these effects did not reach full significance, to obtain a more sensitive measure, we computed *signed confidence ratings* in which the ratings for correct trials were set to [0–100] and those for incorrect trials were negated to [–100–0]. More positive signed confidence ratings reflect more accurate and confident responding; zero reflects chance responding. Consistent with the accuracy and RT measures, the ratings in the related condition were significantly higher for the related (67.1) than for the neutral dimensions (37.0),  $t(11) = 1.82$ ,  $p < .05$ .

We also compared learning of the neutral dimensions across conditions. We found that the neutral dimensions were learned no worse in the related condition as compared to the neutral condition ( $ps > .20$  on all measures), a result consistent with previous studies showing that prior knowledge helps learning without hurting learning of knowledge-unrelated information (e.g., Kaplan & Murphy, 2000). In

the General Discussion we will consider reasons for this lack of difference in the learning of neutral features.

## Discussion

**Experiment 1** replicated standard results in thematic category learning. First, learning occurred in fewer blocks with fewer total errors in the related than in the unrelated condition. Second, single-feature tests showed better learning of the related dimensions than the neutral ones (marginal differences on accuracy and RT, significant differences on signed confidence ratings). Together, these results confirm that the knowledge effect is obtained even when the “prior knowledge” is acquired during an experimental session.

## Experiment 2

The goal of **Experiment 2** was to answer our three main questions about how knowledge affects attention to theme related versus neutral dimensions during learning by replicating Experiments 1’s related condition with an eyetracker. Because the purpose of the unrelated condition was only to confirm that our new materials induced standard knowledge effects, that condition was omitted in **Experiment 2**.

## Method

### Materials

The materials were the same as in **Experiment 1**.

### Participants

Participants were 24 New York University undergraduates who volunteered for \$10. They were randomly assigned in

**Table 3** Single-feature test results from Experiments 1 and 2 (learners only)

	Related condition		Unrelated condition
	Related dimensions	Neutral dimensions	Neutral dimensions
<b>Experiment 1</b>			
Accuracy	0.89	0.71	0.76
Signed confidence rating	67.1	37.0	43.8
RT (seconds)	2.9	4.1	3.5
<b>Experiment 2</b>			
Accuracy	0.91	0.70	
Signed confidence rating	73.6	29.1	
RT (seconds)	2.6	4.0	

equal numbers to one of the four assignments of features to categories and to one of three assignments of related/neutral dimensions.

### Procedure

The procedure was the same as in [Experiment 1](#), with a few additional steps for eyetracking during category learning. Participants were first fitted and calibrated to the eyetracker (SMI system sampling left-eye at 250 Hz). Each trial began with a drift correction that compensated for small movements of the eyetracker on the participant's head. To ensure participants' use of focal vision to obtain feature information, we used a *gaze-contingent display* in which all areas of the screen were blurred except for a circular area around their current point of fixation. After each classification response, auditory feedback was provided, and the whole exemplar was blurred but remained on the monitor for 4 s after the response. Following two practice trials, each training block randomly presented 14 exemplars. The experiment lasted about an hour.

### Eyetracking measures

The eyetracker yields, for each trial, a stream of fixations and their corresponding x-y screen locations and durations. We defined six circular *areas of interest* (AOIs) that encompass the features displayed on the screen<sup>2</sup>. Using the fixations that occurred within the AOIs and before the classification response, we computed four measures in each trial.

The first is the *number of dimensions observed* in each trial. This was a binary measure for each dimension (i.e., observed or unobserved), and thus it ranged from 0 to 4 for the related dimensions and from 0 to 2 for the neutral ones. The second, *fixation probability* (0–1), indicates the probability that a related or a neutral dimension was fixated in a trial. It was computed by dividing the number of dimensions observed by 4 for the related and by 2 for the neutral dimensions. The third, *proportion fixation number* (0–1), was computed by dividing the number of fixations to the related dimensions by total number of fixations to all dimensions. The fourth, *proportion fixation time* (0–1), was computed by dividing the fixation time to the related dimensions by total fixation time to all dimensions. The proportion measures were compared against 0.67 (= 4/6) to reflect the different number of related and neutral dimensions.

<sup>2</sup> For the 20 subjects who learned, 81% of the fixations during training were to those AOIs and the rest were to other areas of the screen.

## Results

### Basic learning results

Once again, participants were very accurate (0.97) in the tests during knowledge acquisition. No participant committed more than a total of seven errors; 12 committed no errors. During training, 20 (of 24) participants reached the learning criterion of two consecutive errorless blocks (6.5 blocks; cf. 5.5 in [Experiment 1](#)) while committing an average of 10.60 errors (cf. 8.67 in [Experiment 1](#)). Classification RTs decreased with more trials but on average they were 0.9 s slower than the related learners in [Experiment 1](#) (possibly because subjects wore eyetracker and were unable to use peripheral vision to obtain feature information, i.e., gaze-contingent display).

### Single-feature test

Table 3 presents single-feature test results. Consistent with [Experiment 1](#), learners were more accurate on the related dimensions (0.91) than the neutral ones (0.70),  $t(19) = 3.31$ ,  $p < .01$ . Signed confidence ratings were higher for the related dimensions (73.6) than for the neutral ones (29.1),  $t(19) = 5.75$ ,  $p < .001$ . Finally, related features were classified faster (2.6 s) than the neutral ones (4.0 s),  $t(19) = 2.92$ ,  $p < .01$ .

### Eye fixations

Figure 3 shows the eye-fixations results averaged over the 20 learners during the course of training. Each data point represents fixations averaged over a subblock of four trials. In addition, we assumed that learners' eye movements would have been identical to those in their last subblock had they continued classifying for the full 15 blocks (so that every subject contributes to each data point).

Figure 3a shows that learners initially observed about three of four related dimensions and gradually increased fixations to those dimensions over the course of training. In contrast, they initially observed about 1.5 of the two neutral dimensions and those fixations gradually decreased. Figure 3b presents fixation probabilities that equate different number of related and neutral dimensions. The figure indicates that learners fixated the two types of dimensions with about equal probability at the start of training but became more (less) likely to fixate the related (neutral) dimensions. By the end of training, they were more than twice as likely to fixate the related dimensions than the neutral ones (0.89 vs. 0.35).

A  $2 \times 2$  within-subjects ANOVA was conducted on the fixation probabilities in Fig. 3b with dimension type (related vs. neutral) and subblock (first vs. last) as factors.

**Fig. 3** Eye-fixations results from Experiment 2. (a) Number of related/neutral dimensions fixated. (b) Probability of fixation to the related/neutral dimensions. (c) Proportion fixation number/time. The 0.67 (= 4/6) line reflects fixation proportion in favor of neither dimension types

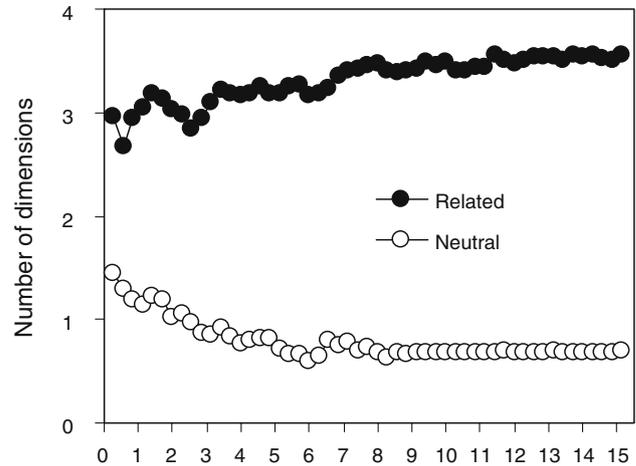
There was a main effect of dimension type,  $F(1, 19) = 17.61$ ,  $MSE = .087$ ,  $p < .001$ , confirming the greater chance of fixating the related dimensions. There was no main effect of block ( $p > .10$ ) but a significant interaction between dimension type and subblock,  $F(1, 19) = 26.53$ ,  $MSE = .051$ ,  $p < .001$ , confirmed the increase (decrease) in fixating the related (neutral) dimensions. Paired  $t$ -tests in each subblock (Fig. 3b) revealed that learners were more likely to fixate the related dimensions than the neutral ones in all subblocks,  $p$ 's  $< .05$ , except subblocks 1, 2, 3, and 6, indicating that they did not have a preference for attending to the related dimensions until after almost a full training block.

These results are further supported by the more sensitive proportion measures in Fig. 3c. Because there were four related and two neutral dimensions, a value of 0.67 (= 4/6) reflects a bias toward neither dimension type. The figure shows that both proportion fixation number and time start off around 0.67 and then shift in favor of the related dimensions.  $T$ -tests comparing the first and last subblock confirmed increase in both proportions,  $p$ 's  $< 0.001$ . In addition, both proportions were greater than 0.67 in all subblocks,  $p$ 's  $< .05$ , except subblocks 1, 2, 5, 6, 7, and 8. These results are consistent with the fixation probabilities in Fig. 3b indicating that learners' preference for the related dimensions emerged only after the observation of category members (and the receipt of error feedback).

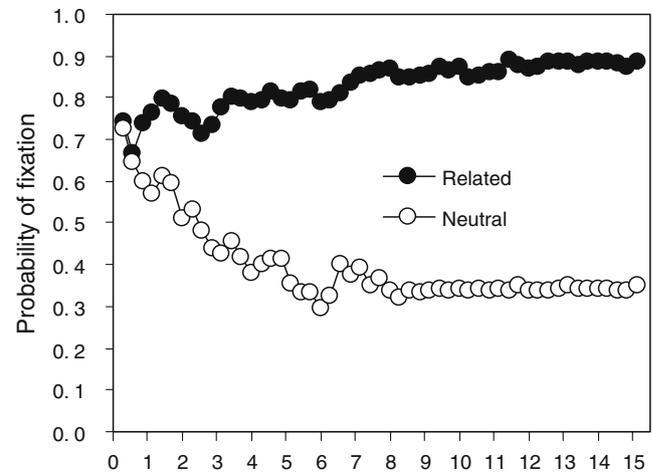
#### Backward learning curves

The previous analyses indicate learners' gradual shift in attention to related dimensions during the course of training. We also asked how that shift relates to error, that is, whether negative feedback is required for shifts in attention. To answer this question, we created backward learning curves (Fig. 4) by translating each subject's trial numbers so that their last error occurred on trial 0 (and thus subblock 0 always included the last error trial). Figure 4 includes ~10 blocks before the last error and ~3 blocks after. (Trials after the last error include those from the last two error-free blocks plus those correct trials from the end of the previous block.) We padded out learner's eye movements in their first and last subblocks to the left and right of Fig. 4, respectively, so that every subject contributes to each data point.

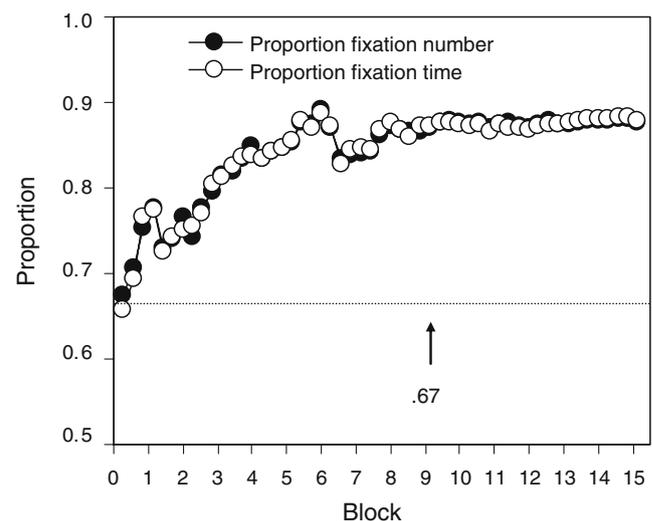
#### A. Number of dimensions fixated



#### B. Probability of fixation



#### C. Proportion fixation number/time



**Fig. 4** Backward learning curves from Experiment 2. (a) Number of related/neutral dimensions fixated. (b) Probability of fixation to the related/neutral dimensions. (c) Proportion fixation number/time

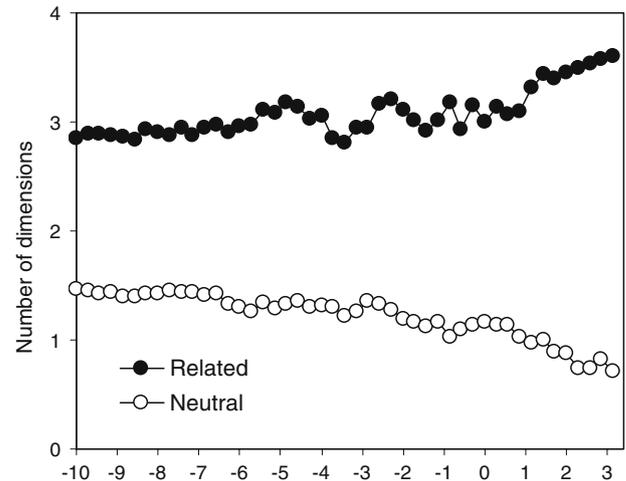
Figure 4a presents the number of dimensions observed in each subblock. Of greater interest are Figs. 4b and c which present fixation probabilities and proportion fixation number and time, respectively. First, consider the eye fixations that occurred before the last error (i.e., on negative blocks). Figure 4b indicates that the two probabilities were initially indistinguishable but gradually increased for the related dimensions and decreased for the neutral ones. Likewise, Fig. 4c shows that both proportions were initially around 0.67 but gradually increased in favor of the related dimensions. Paired *t*-tests in each subblock that compared fixation probabilities between related versus neutral dimensions (Fig. 4b) revealed that learners were more likely to fixate the related dimensions from the subblock indicated by an arrow,  $p$ 's < .05. Both the proportion fixation number and time measures (Fig. 4c) were also significantly greater than 0.67 from the same subblock indicated by an arrow,  $p$ 's < .05. These results indicate that learners began to direct their attention to the related dimensions a few blocks before the last error.

Next, consider the eye fixations after the last error. Both Figs. 4b and c indicate that the shift in attention continued after the last error, that is, despite the absence of negative feedback. After the last error, fixation probabilities for the related dimensions rose from 0.78 to 0.90, and those for the neutral dimensions dropped from 0.57 to 0.36. A  $2 \times 11$  within-subjects ANOVA was conducted on the fixation probabilities (Fig. 4b) with dimension type (related vs. neutral) and subblock (1 to 11) as factors. There was a main effect of dimension type,  $F(1, 19) = 22.92$ ,  $MSE = .717$ ,  $p < .001$ , confirming the greater chance of fixating related dimensions. There was no effect of subblock ( $F < 1$ ), but a significant interaction between dimension type and subblock,  $F(10, 190) = 5.58$ ,  $MSE = .027$ ,  $p < .001$ , indicated the increase (decrease) in fixating the related (neutral) dimensions. Considering the two types of dimensions separately, fixation probabilities increased from subblock 1 to subblock 11 for the related dimensions and decreased for the neutral dimensions,  $p$ 's < .05. In Fig. 4c, the two proportions also showed a reliable increase from the subblock 1 to 11,  $p$ 's < .01.

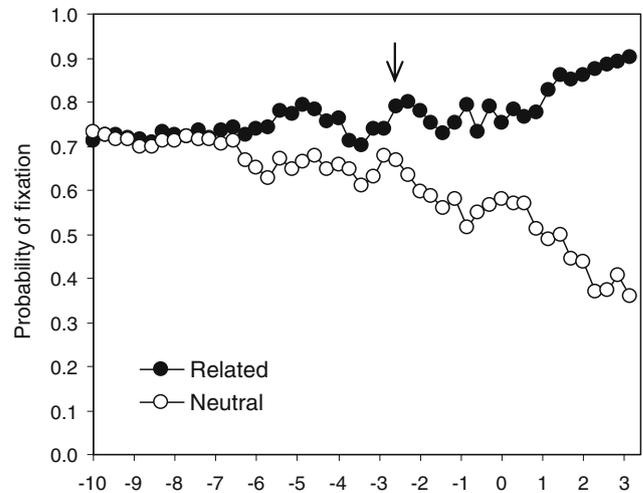
*Individual variation*

We asked whether the patterns of eye fixations in Figs. 3 and 4 were manifested consistently by all learners. We identified six learners whose eye movements were similar

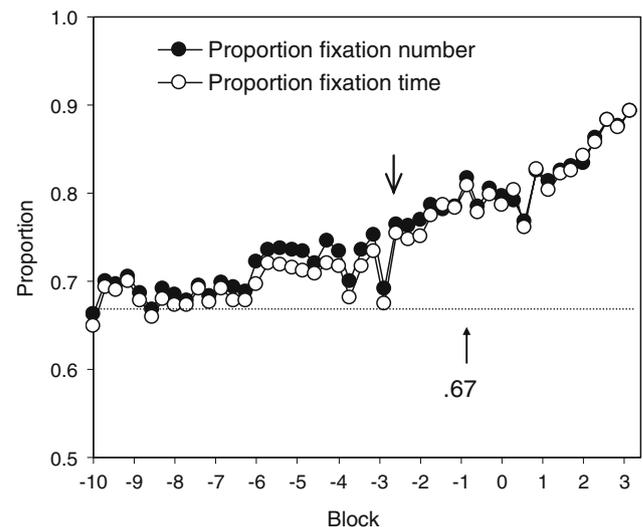
**A. Number of dimensions fixated**



**B. Probability of fixation**



**C. Proportion fixation number/time**



to one another but distinct from the group average. As compared to other subjects, this group fixated all six dimensions at the start of training, showed at best only a small preference for the related dimensions before committing their last error, and learned the categories very quickly (average number of total errors was 3.0). Although we do not wish to over-interpret the performance of a small number of subjects, we think it likely that these individuals recognized the category themes before the start of classification training (i.e., during the knowledge acquisition phase), and thus learned the categories quickly because they only needed to learn which theme went with which category label. Nevertheless, just as with the group data in Fig. 4, these subjects shifted attention away from neutral dimensions after their last error. Note that the gradual shift of attention away from neutral dimensions implied by Figs. 3 and 4 was not an artifact of averaging over subjects; examination of the eye movements of individual subjects revealed that those shifts were indeed gradual for the large majority of subjects.

#### *Relating eye fixations to feature learning*

Finally, we investigated whether more eye fixations to a dimension during training resulted in better learning of that dimension. Accordingly, we performed a simple regression for each participant where a proportion fixation number was used to predict the signed confidence ratings from the single-feature test (average  $R^2 = .45$ ). The weight assigned to the proportion fixation number averaged over subjects (298.0) was significantly greater than 0,  $t(19) = 4.33$ ,  $p < .001$ , indicating that for each 0.10 increase in the proportion fixation number, the signed confidence rating increased by 29.8. That is, more fixations to a dimension led to better learning of that dimension. In addition, the mean intercept (9.1) did not differ from 0 ( $t < 1$ ), indicating that no learning of a dimension occurred if it was never fixated. Similar results were obtained when proportion fixation time was used as a predictor.

We also asked whether the better learning of related dimensions could be explained solely by the greater number of fixations they received. We conducted additional per subject regressions in which a variable coding whether the dimension was related (1) or neutral (0) was added to the proportion fixation number as a predictor. In this analysis (average  $R^2 = .70$ ), both proportion fixation number (209.3),  $t(19) = 2.66$ ,  $p < .05$ , and dimension type (28.8),  $t(19) = 2.45$ ,  $p < .01$ , were significant predictors. That is, the signed rating increased by 20.9 for each 0.10 increase in proportion fixation number and by 28.8 when a dimension was related rather than neutral. Thus, the

better learning of related versus neutral dimensions was partially but not fully mediated by the extra attention they received. Similar results were obtained when proportion fixation time was used as a predictor.

#### Discussion

Experiment 2 answered our three main questions regarding the effects of knowledge on attention. First, eye fixations showed that prior knowledge indeed affects what category information is attended, as learners ended up allocating more attention to related dimensions than neutral ones. Second, learners showed no initial tendency to fixate related dimensions. Rather, they gradually shifted attention to related dimensions during the course of training. Third, this shift in attention continued after the classification problem was solved, that is, in the absence of negative feedback. Finally, although eye fixations during training were a significant predictor of feature learning at test, they did not fully mediate the better learning of the related dimensions.

#### General discussion

This article has addressed how prior knowledge affects attention to features of to-be-learned categories. Although numerous investigators have considered the possibility that knowledge affects attention, without direct evidence such proposals have remained speculative. We now discuss the implications our results have for the three questions we posed in the Introduction and for models of knowledge-based category learning. The final section relates knowledge's effect on attention to others it has on category learning.

#### An effect of prior knowledge on attention

Our first question was whether in fact knowledge induces any change to selective attention. Earlier we noted how knowledge might exert its effect solely through how category information is encoded or by allowing classification to become an act of inference in which people reason from features to category membership. Instead, we found that a preference for the related versus neutral features emerged in learners after about a block training. By the end of training, these subjects were more than twice as likely to fixate the related features. This finding is the first direct confirmation of the proposal that knowledge directs attention to knowledge-relevant information (e.g., Heit & Bott, 2000; Kruschke, 1993; Murphy & Medin, 1985; Murphy & Allopenna, 1994; Pazzani, 1991).

These results have implications for models of knowledge-based learning. Past studies have shown that both KRES and Baywatch correctly predict the faster learning in the presence of knowledge and the better learning of related features versus neutral ones (Heit & Bott, 2000; Rehder & Murphy, 2003). However, these models predict poorer learning of neutral features not because of reduced attention but because of various forms of cue competition that arise from error-driven learning. The processes used to explain the better learning of related features are analogous to those used by the Rescorla-Wagner learning rule to account for the phenomenon of *overshadowing* in animal learning (Kamin, 1969; Rescorla & Wagner, 1972); the faster learning of the related features results in error being reduced more rapidly, which slows the formation of associations between the neutral features and the category label.

However, it is well known that many standard effects of cue competition can arise from not only error-driven learning but also from attentional mechanisms (Kruschke, 2001, 2003; Kruschke & Blair, 2000; Kruschke et al., 1999; Macintosh 1975; Sutherland & Mackintosh, 1971). The present results show that attentional effects occur in knowledge-based category learning as well—because learners attend the neutral features less often, they will be learned less well than the related ones. Thus, neutral features are likely at a double disadvantage in learning, as both reduced error and reduced attention result in them being more weakly associated with the category label. (Later we will identify additional processes that may compensate for this reduced learning of neutral features).

The second implication that our eye movement results have for models concerns how items ended up being categorized at the end of training. That neutral features are less strongly associated with the category label of course means that they contributed less to learners' accurate classification performance. But on top of that, at the end of training the neutral features were fixated less often than the related ones. In other words, the neutral features were at a double disadvantage in classification as well—they provided a relatively weak source of evidence for category membership that was largely ignored anyway.

Other studies provide evidence suggestive of an effect of knowledge on attention during classification. For example, Lin and Murphy (1997) found that subjects were more sensitive to features of novel artifacts that were relevant to the artifacts' stated purpose as compared to unrelated features even when the stimuli were presented for only 50 ms (and followed by a mask) (also see Luhmann, Ahn, & Palmeri, 2006; Palmeri & Blalock, 2000). But these results may have been due either to the greater weight on knowledge related features during classification or to the

greater attention those features received during training. In contrast, the present study provides unambiguous evidence for an effect of knowledge on attention during both initial learning and subsequent classification.

That models like Baywatch and KRES ignore the fact that neutral features receive fewer attentional resources means that they have mistakenly attributed the poorer learning of those features solely to effects of error-driven cue competition. And, doing so means that they have mistakenly attributed neutral features' limited influence on final classification performance solely to their poorer learning. In other words, in the absence of attentional mechanisms that direct resources toward knowledge relevant information, Baywatch and KRES mischaracterize the effects of prior knowledge on how features are learned and their ultimate role in classification performance.

#### Knowledge selection and construction in response to observed category members

The second question we asked concerned the time course of the effect of knowledge on attention. Earlier we reviewed proposals suggesting that the impact of knowledge consisted solely of selecting which sources of information are considered at the start of learning (e.g., Heit, 1995; Kruschke, 1993; Pazzani, 1991). Instead, we found that learners' preference for the knowledge-related dimensions increased as training progressed. Indeed, at the start of training learners were no more likely to fixate related dimensions than neutral ones.

It is important to note that the studies of Heit (1995) and Pazzani (1991) that documented early effects of knowledge differed from the present one in a crucial way, namely, that the category labels being predicted (attending parties and a balloon inflating, respectively) were already familiar to subjects and so provided, right from the start of learning, a cue to what prior knowledge was likely to be relevant (also see Wisniewski & Medin, 1994; Wisniewski, 1995). But although category labels are sometimes familiar, the labels of most new categories are themselves new (e.g., the label “iPod” was initially as opaque to you as “Kez” was to our subjects). In these cases, prior knowledge must enter instead through the semantic associates of features of category members. But because there are many features and each have many associations, determining which semantic representations are relevant to the current learning problem will often occur only after several category members are observed.

Heit and Bott (2000) labeled the process by which observations activate relevant semantic representations as a “knowledge selection,” and, like us, emphasize that

many observations may be required before relevant knowledge is identified. For example, although our subjects may have tried to make use of feature descriptions with phrases such as “slippery ground”, “low temperature”, and “hard soil”, it may not have been immediately obvious how those phrases were related to each other. However, repeated presentation of the features (and repeated recall of the feature descriptions) eventually allowed them to triangulate onto what these features had in common: that the ground was slippery because it was *icy* (rather than merely wet), that the soil was hard because it was *frozen* (rather than just highly compacted), and both of these things were true because the temperature was not just “low” but *below freezing*. And of course learners only began to direct attention toward theme-relevant features after they started to realize that the ant was adapted to a cold, icy environment. In contrast, models like Baywatch and KRES, assume that knowledge is in place from the start of training rather than being constructed in response to observed category members.

Of course, how quickly relevant knowledge is activated will depend on how effectively observed category members serve as retrieval cues for that knowledge. Although it appeared to start only at the end of the first block of training in our study, it may occur more quickly with other sorts of materials. For example, whereas Heit and Bott (2000) found that subjects were no more accurate in classifying related features after the first training block when learning church and office buildings (labeled “Does” and “Lees”, respectively), they were when the categories were types of tractors and race cars (see Kaplan & Murphy, 2000, for another example of knowledge effects that emerge early in learning). Even in the present study, the delayed effect of knowledge might have been due to the fact that it was provided as part of the experiment and thus may not have been encoded as strongly (or retrieved as readily) as real-world semantic representations. But regardless of when effects of knowledge begin, the import of the present study is in demonstrating that those effects are not limited to merely preselecting which stimulus dimensions are attended.

#### The unnecessary role of error in attending knowledge-relevant information

Our third question was whether error feedback is required to mediate shifts in attention to knowledge-relevant information. We noted that all current accounts of how attention changes during learning are based on error. For example, ALCOVE predicts gradual shifts in attention to stimulus dimensions that reduce error (Kruschke, 1992).

Hypothesis testing models also assume that attention shifts between dimensions when classification errors result in the rejection of old rules (Nosofsky et al., 1994; also see Kruschke & Johansen, 1999). But, contra this account, attention continued to shift to the related dimensions even after subjects learned to classify all items, that is, in the absence of negative feedback. Thus, error is not a necessary condition for knowledge-induced changes in attention.

We propose two possible explanations for shifts in attention in the absence of negative feedback. The first is the processes of theme discovery we have described, that is, through the activation of semantic representations common to several category features. In our experiments, merely observing features of to-be-classified stimuli may have been sufficient for learners to activate related representations, enabling the discovery of the tundra and desert themes. In fact, the extensive literature documenting knowledge effects in unsupervised category construction tasks suggests that the discovery of category themes can occur in the absence of any sort of feedback (Kaplan & Murphy, 1999; Medin et al., 1987; Spalding & Murphy, 1996). That learners in *Experiment 2* shifted attention to theme-related dimensions after errors ceased suggests that spontaneous theme elaboration can also occur during supervised classification learning. For example, the relatedness of the four knowledge-related dimensions need not have occurred at the same time—learners may have noticed the thematic relationship between two or three of the related dimensions before committing their final error and discovered the others afterwards.

A second reason that attention might shift without error is that it is likely that our cognitive systems are trying not only to increase *accuracy* but also decrease *response time*—all else being equal, a faster classification is more adaptive than a slower one. Indeed, the response times of learners in *Experiment 2* decreased from 8.7 s at the point of their last error to 4.4 s at the end of training. One way that latency can be decreased is by gathering less information in preparation of a decision, and of course to maintain accuracy low quality sources of information should be discarded before higher-quality ones. On this account, the need for speed led our learners to recognize that fewer dimensions were needed for accurate performance—in our category structure only three of six dimensions were required for perfect classification. Given a choice, the poorly learned neural dimensions were the first to go (see Nelson & Cottrell, 2007, for one computational implementation of this idea).

Moreover, studies have found shifts in attention in the absence of even positive feedback. For example, Blair et al. (2009a) found that learners continued to optimize attention even after a criterion of 24 correct trials was reached and

feedback stopped altogether. Of course, attentional shifts in the absence of feedback pose problems for all category learning models that tie attentional learning to error-driven mechanisms (e.g., Kruschke, 1992).

These possibilities suggest that knowledge-induced attention shifts can be both a *cause* and an *effect* of learning. On the one hand, prior knowledge can direct attention to information needed for learning. But attention shifts can also reflect learning that has already occurred, as when less valuable sources of information are bypassed in order to respond more rapidly.

#### Attention versus encoding, inference, and interpretation in knowledge-based category learning

Finally, we also asked whether the effect of prior knowledge on category learning can be understood as being fully mediated by its influence on attention. The answer is that it cannot. Although eye movements to knowledge-relevant features were indeed predictive of their greater learning, we found that those features were learned better than neutral ones even controlling for eye fixations. We now review several mechanisms via which knowledge may influence learning besides attention.

First, there is evidence for the sort of encoding processes enabled by prior knowledge we have mentioned. Although we found a learning advantage for related over neutral features in the presence of prior knowledge in [Experiment 1](#), we also found that these neutral features were learned *no worse* than neutral features in the unrelated control condition. These results replicate those by Kaplan and Murphy (2000) who also found no evidence that neutral features were learned worse in the presence of knowledge. These results are surprising given the standard error-driving learning accounts responsible for cue competition we have reviewed. They are doubly surprising in the light of our results that neutral features are also attended less often. We believe that these results have arisen because knowledge affects not just how features are attended but also how they are processed and encoded. For example, Kaplan and Murphy also found evidence that learners attempted to assimilate the supposedly neutral features to the categories' themes (also see Heit, Bott & Briggs, 2004). In other words, when knowledge is present, learning is not a zero-sum game. Instead, it provides the knowledge structures that promote the effective encoding of many sources of information, even those that are only peripherally related to that knowledge.

Second, we have also mentioned how in many cases prior knowledge allows classification to become an act of

inference. For example, in Heit and Bott's (2000) study, subjects' mental representation of Doe buildings probably included the fact that they were "church like," suggesting that during classification they used observed features to infer the concept "church" and from that the concept label "Doe" (this inferential process is explicit in their Baywatch model). Consistent with this interpretation, Heit and Bott found that learners classified a feature as "Doe" even if it was never observed during training so long as their prior knowledge indicated that it was typical of churches. Other studies provide evidence of the inferences in service of classification that knowledge supports (Rehder & Kim, 2009; Rehder & Ross, 2001). Indeed, recall Murphy and Medin's (1985) example of classifying a partygoer who jumps into a pool as *drunk*—one reasons from aberrant behavior to its underlying cause even if one has never before observed drunken swimming. Clearly, classification performance on novel items cannot be explained in terms of how those items were attended and encoded during training.

Finally, prior knowledge can also influence how stimuli are interpreted in the first place. For example, like the studies of Pazzani (1991) and Heit (1995) we have reviewed, Wisniewski and Medin (1994) used familiar category labels but, unlike those studies, used ambiguous pictorial stimuli. They found that the category labels influenced how the pictures were interpreted. For example, in the same picture, a character was interpreted either as "dancing" when subjects were told the pictures were drawn by creative versus noncreative kids (dancing was taken to be a sign of creativity) or as "climbing in a playground" when told they were drawn by children who lived in cities versus farms (playgrounds are in cities but not on farms). For present purposes, the important point is that these differences in interpretation do not necessarily require differences in *attention*—the same stimulus can be attended equally, but interpreted differently.

In summary, although we think that attention is an important vehicle by which knowledge influences category learning, it also exerts its influence through other means, including how stimulus items are encoded, the inferential processes it supports, and how features are interpreted. Nevertheless, models such as Baywatch and KRES will remain incomplete until they include mechanisms by which knowledge activated in response to observed category members and which then directs attention to toward knowledge-relevant information.

**Acknowledgments** This material is based upon work supported by the National Science Foundation under Grant No. 0545298. We thank Gregory L. Murphy, Harlan D. Harris, and anonymous reviewers for their comments on an earlier version of this manuscript.

## Appendix

### Materials for Experiments 1 and 2

The features and their associated knowledge are presented. There were 12 features, two features for each of the six dimensions. For each feature, three types of knowledge (i.e., Tundra, Desert, & Neutral) were invented resulting in 36 descriptions in total.

Antenna (0)



- 0<sub>t</sub>: The ants need to photosynthesize amino acids to sustain their life. Because the daytime is short, they use this fan type of antennae to maximize the surface area exposed to the sunlight.
- 0<sub>d</sub>: Because the air is hot and dry, the ants are vulnerable to dehydration. To maintain hydration, the ants use this fan type of antennae to absorb water vapor from the air.
- 0<sub>n</sub>: The ants use sunlight and moonlight to orient themselves. This fan-like antennae allows to absorb enough light for purpose of orientation.

Antenna (1)



- 1<sub>t</sub>: Because the temperature is very low, parts of ants' eyes (e.g., cornea, iris, pupil) often freeze and the ants become blind. When that happens, this thread type of flexible antennae is used to detect close objects.
- 1<sub>d</sub>: Because the temperature is very high, the ants need to dissipate excess body heat. This thread type of antennae promotes heat dissipation.
- 1<sub>n</sub>: The ants become blind at night and use this thread type of antennae to detect close objects.

Foot (0)



- 0<sub>t</sub>: Because the ground surface is extremely cold, the ants conserve body heat by switching the toe that comes into contact with the ground in each step.
- 0<sub>d</sub>: Because the ground surface is extremely hot, the ants switch the toe that comes into contact with the ground in each step to avoid burning.
- 0<sub>n</sub>: The ants communicate through chemicals called pheromones. Each of these three toes releases a unique chemical to convey different messages.

Foot (1)



- 1<sub>t</sub>: Because the ground surface is slippery, the ants need to have wide feet to maintain their footing.
- 1<sub>d</sub>: Because of the sandy soil, the ants have wide feet that prevent them from sinking below the surface.
- 1<sub>n</sub>: The ants protect themselves from enemies approaching from behind by kicking with these sharp protrusions.

Forearm (0)



- 0<sub>i</sub>: Because the ground is slippery in the ants' environment, the forearm with thorn-like protrusions helps the ants to move without slipping.
- 0<sub>d</sub>: Because the ants' prey (e.g., fleas) hide in sand, the ants use this type of forearm to sweep the sand and detect the prey.
- 0<sub>n</sub>: When the ants engage in a fight, they use this saw-like forearm to tear the enemy apart.

Forearm (1)



- 1<sub>i</sub>: Because of frequent blizzards, the ants need to anchor themselves during high winds. This type of forearm allows the ant to hold its position.
- 1<sub>d</sub>: Because of strong direct sunlight, the ants dig deep into the ground in order to be cool when they rest. This type of forearm helps the ants to do the job easier and faster.
- 1<sub>n</sub>: The ants sometimes plunder other ants' colony of eggs. This hook-type of forearm is useful in digging in search of the eggs.

Mouth (0)



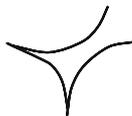
- 0<sub>i</sub>: Because sources of food are frozen and tough, the ants mash and grind them using the upper and lower parts of the mouth before swallowing.
- 0<sub>d</sub>: Because sources of food are covered with sand, they need to be cleared before swallowing. The inner surface of the ants' mouth has short but stiff hairs that filter out these impurities.
- 0<sub>n</sub>: The ants are herbivorous. This long mouth is used to grind tough fibroid materials in plants before they are swallowed.

Mouth (1)



- 1<sub>i</sub>: Because the ground is frozen, the ants need to cut and break tough soil in search of their food. This type of mouth with sharp incisors serves this function.
- 1<sub>d</sub>: Because the air is dry and the sunlight is strong, food dries out quickly. The ants hold their food in the cavity of their mouth on the way to their nest so that the food does not become dry.
- 1<sub>n</sub>: The ants need to transport food to their colony. This mouth allows them to hold the food in the cavity of their mouth until they arrive at their colony.

Tail (0)



- 0<sub>i</sub>: Because water tends to exist in a frozen state, the ants acquire water by collecting dew drops with this trumpet-shaped tail early in every morning.
- 0<sub>d</sub>: Because the air is dry and water is scarce, the ants need to collect water whenever possible. This trumpet-shaped tail is used to collect rain during the rare rainstorm.
- 0<sub>n</sub>: The ants lay a large number of eggs at a time. This trumpet-shaped tail allows the ants to deliver a large number of eggs.

## Tail (1)



$1_i$ : Because water exists in a frozen state on the surface, the ants have to find melted water beneath the surface. This type of tail probes into places where melted water might exist.

$1_d$ : Because the air is dry and water is scarce, the ants store extra water in the humps of their tail.

$1_n$ : The ants feed proteins stored in the humps to their larvae using the sharp nozzle in the end of tail.

## Wing (0)



$0_i$ : Because of low temperature, the wing ends become tattered. The gray area in the wings promotes cell growth by which the damaged wing ends are replaced with new ones.

$0_d$ : Because of high temperature, the wing ends become tattered. The gray area in the wings promotes cell growth by which the damaged wing ends are replaced with new ones.

$0_n$ : The ants have red spots in the wing ends. The color becomes brighter in the mating season by the hormones produced in the gray area.

## Wing (1)



$1_i$ : Because hailstones hit the ants' wings and tear them apart, the ends of their wings have thick veins that protect them from being split.

$1_d$ : Because sandstorms tear the ants' wings apart, the ends of their wings have thick veins that protect them from being split.

$1_n$ : While flying, the ants control their rapid changes in direction by adjusting the fore- and rear-flaps in each wing.

## References

- Blair, M., Watson, M. R., & Meier, K. M. (2009a). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*, 330–336.
- Blair, M., Watson, M. R., Walshe, R. C., & Maj, F. (2009b). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1196–1206.
- Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorical word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 172–190.
- Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, *34*, 686–708.
- Harris, H. D., & Rehder, B. (2006). Modeling category learning with exemplars and prior knowledge. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1440–1445). Mahwah, NJ: Erlbaum.
- Heit, E. (1995). Belief revision in models of category learning. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual conference of the cognitive science society* (pp. 176–181). Mahwah, NJ: Erlbaum.
- Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of learning and motivation* (pp. 163–199). San Diego: Academic Press.
- Heit, E., Briggs, J., & Bott, L. (2004). Modeling the effects of prior knowledge on learning incongruent features of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1065–1081.
- Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 151–182). Hillsdale, NJ: Erlbaum.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In R. M. Church & B. A. Campbell (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century Crofts.
- Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, *27*, 699–712.
- Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 829–846.
- Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review*, *88*, 197–226.
- Krascum, R. M., & Andrews, S. (1998). The effects of theories on children's acquisition of family resemblance categories. *Child Development*, *69*, 333–346.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

- Kruschke, J. K. (1993). Three principles for models of category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *Categorization by humans and machines: The psychology of learning and motivation* (Vol. 29, pp. 57–90). San Diego: Academic Press.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.
- Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*, 171–175.
- Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*, 636–645.
- Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1083–1119.
- Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, *42*, 267–316.
- Lin, E. L., & Murphy, G. L. (1997). The effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1153–1169.
- Luhmann, C. C., Ahn, W., & Palmeri, T. J. (2006). Theory-based categorization under speeded conditions. *Memory & Cognition*, *34*, 1102–1111.
- Macintosh, N. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.
- Maddox, W. T. (2002). Learning and attention in multidimensional identification and categorization: Separating low-level perceptual processes and high-level decisional processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 99–115.
- Maddox, W. T., & Dodd, J. L. (2003). Separating perceptual and decisional attention processes in the identification and categorization of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 467–480.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*, 242–279.
- Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904–919.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Nelson, J. D., & Cottrell, G. W. (2007). Probabilistic model of eye movements in concept formation. *Neurocomputing: An International Journal*, *70*, 2256–2272.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 149–167). Hillsdale, NJ: Erlbaum.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.
- Palmeri, T. J., & Blalock, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition*, *77*, B45–B47.
- Pazzani, M. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 416–432.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
- Rehder, B., Colner, R. M., & Hoffman, A. B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language*, *60*, 394–419.
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 811–829.
- Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition*, *37*, 715–729.
- Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review*, *10*, 759–784.
- Rehder, B., & Ross, B. H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1261–1275.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. R. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, (13, Whole No. 517).
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167–196.
- Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 525–538.
- Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
- Watson, M., & Blair, M. (2008). *Attentional allocation during feedback: Eyetracking adventures on the other side of the response*. Paper presented at the proceedings of the 28th annual conference of the cognitive science conference, Vancouver, Canada.
- Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 449–468.
- Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, *18*, 221–281.